

youNa

Wie aus einem regelbasierten Chatbot ein empathisches, LLM-gestütztes System wurde – für stärkende Unterstützung im Umgang mit Rassismuserfahrungen.

Einführung

Die Case Study beleuchtet die Weiterentwicklung von YANA & zeigt die wichtigsten Meilensteine des Projekts. Sie fokussiert sich darauf, welche technischen und inhaltlichen Entscheidungen den Chatbot verbessert haben. Im Mittelpunkt stehen die Entwicklungsschritte, Learnings und der Einfluss auf die Nutzer:innen.

Auch wenn es durch mediale Aufmerksamkeit oft den Anschein erweckt, als wäre Rassismus mittlerweile ein zentrales Thema, mit dem sich unsere Gesellschaft kritisch auseinandersetzt, sieht die Realität für BIPOC in Deutschland noch immer anders aus. Strukturelle Benachteiligung und Diskriminierung sind in vielen Bereichen des alltäglichen Lebens allgegenwärtig – sei es in Behörden, im Bildungssystem oder im Arbeitsmarkt. Und obwohl zahlreiche Menschen betroffen sind, gibt es nach wie vor keine flächendeckende Unterstützung. Laut Studien haben 22 Prozent der deutschen Bevölkerung bereits rassistisch motivierte Ausgrenzung erfahren – das sind fast 19 Millionen Menschen. Demgegenüber stehen durchschnittlich nur eine Vollzeit-Beratungsstelle pro 920.000 Einwohner.

Viele erleben ihre Situation als isolierend und stoßen zusätzlich auf Infragestellungen oder Bagatellisierungen ihrer Erlebnisse. Auch die aktuelle politische Lage in Deutschland sowie weltweit kann ein zusätzliches Gefühl der Machtlosigkeit und Unsicherheit verstärken. Es fehlt an niedrigschwelligen Unterstützungsangeboten, die sowohl emotionale Begleitung als auch rechtliche Orientierung bieten.

Aus der Motivation heraus, Menschen mit Rassismuserfahrungen eine verlässliche Ressource an die Hand zu geben, gründete Said Haider 2021 das Projekt YANA, das später in YOUNA umbenannt wurde. Ziel war es, ein Tool zu entwickeln, das nicht nur emotionalen Beistand leistet, sondern auch durch das Melden von Vorfällen zur Sichtbarmachung von Diskriminierung beiträgt.

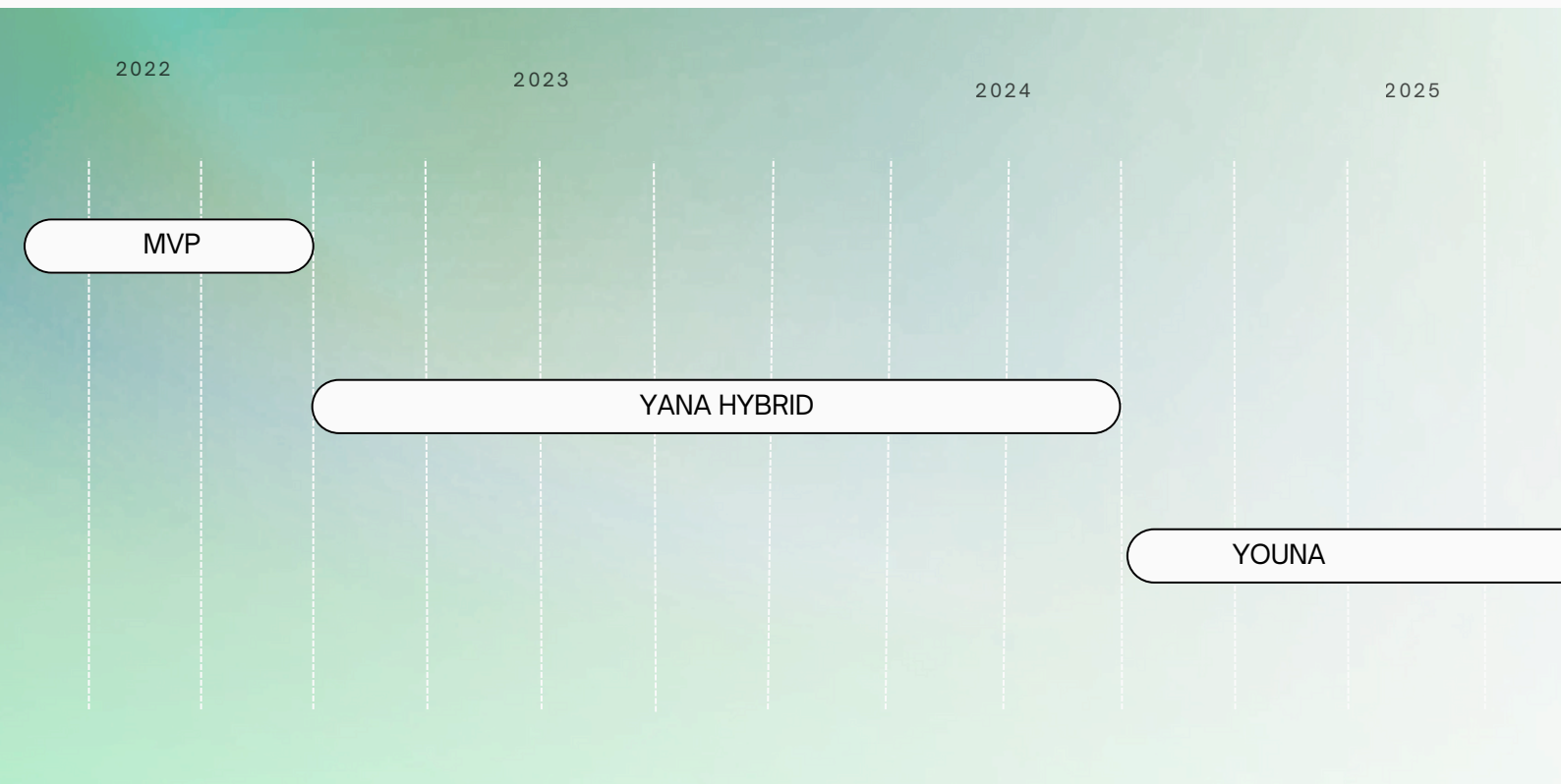
Gemeinsam mit Expert:innen aus den Bereichen Antidiskriminierung und Jura entstand 2022 in Zusammenarbeit mit Parslabs ein erstes MVP – ein regelbasierter Chatbot, der auf Rasa entwickelt wurde. Während Parslabs die technische Umsetzung und die Gestaltung der Benutzeroberfläche übernahm, lag unser Fokus auf der Konzeption der Bot Persona, dem Dialogdesign und der inhaltlichen Aufbereitung.

Da es für einen solchen Use Case keine vergleichbaren Referenzprojekte gab, bestand eine der größten Herausforderungen in der Anfangsphase darin, die Bedürfnisse der Nutzer:innen zu verstehen und die Funktion des Chatbots so zu definieren, dass er gezielt in den wichtigsten Bereichen Unterstützung leisten kann.

Die erste Version von YANA war ein Button-basierter Chatbot, der jedoch schnell an seine Grenzen stieß. Es wurde deutlich, dass dieser regelbasierte Ansatz nicht annähernd die Komplexität abbilden konnte, die notwendig war, um den individuellen Anliegen der Nutzer:innen gerecht zu werden. Der Dialog war zu starr, zu vorgegeben – es fehlte die Möglichkeit, Erfahrungen so zu schildern, wie es gebraucht wurde.

Diese Erkenntnis markierte den Beginn einer Entwicklungsreise, die von kontinuierlichem Testing, Re-Designs und technologischen Weiterentwicklungen geprägt war. Nach einer hybriden Übergangsphase fiel 2024 schließlich die Entscheidung, vollständig auf einen LLM-gestützten Ansatz umzusteigen – ein System, das nun in der Lage ist, kontextsensitive, empathische und bedarfsgerechte Antworten zu liefern.

Development



MVP In der Anfangsphase lag unser Fokus darauf, eine Bot-Persona zu entwickeln, sich durch eine authentische und stimmige Charaktergestaltung auszeichnet. Gleichzeitig war anfangs noch nicht klar, welche Themen im Mittelpunkt stehen sollten und wie die inhaltliche Struktur des Chatbots aussehen könnte. Durch eine Kombination aus User-Befragungen und Expertengesprächen haben wir relevante Informationen gesammelt, analysiert und aufbereitet. Daraus entstand schließlich die folgende thematische Ausrichtung:

YANA sollte nicht nur eine klassische Unterstützung bieten, sondern als ehrenamtliche:r Begleiter:in auftreten – eine Mischung aus beste:r Freund:in und Anwält:in. Ziel war es, eine vertrauensvolle, warme und persönliche Gesprächsatmosphäre zu schaffen, die den Nutzer:innen einen sicheren Raum bietet.

2022 launched the project as a rule-based MVP with Button-Navigation, followed by a hybrid phase 2023, which integrated the first Freitext-Elements. In 2024, the transition to an LLM-supported system for more flexible and empathetic dialogues.

Dabei wurde darauf geachtet, dass die Persona eine komplexe, aber nicht überwältigende Charaktertiefe besitzt. YANA sollte ein Ally sein, eine verständnisvolle Stimme, die durch das zentrale Thema führt: "You Are Not Alone".

Diese Leitidee spiegelte sich auch im Gesprächsverlauf wider, der sich auf drei zentrale Bereiche konzentriert:

⁰¹ Heal

Nutzer:innen, die mit YANA sprechen, suchen oft nach psychologischer Unterstützung, Selbstbildung oder Strategien zur emotionalen Bewältigung. Das Ziel in diesem Bereich war es, hilfreiche Impulse zu bieten, um das Erlebte weniger belastend erscheinen zu lassen und Möglichkeiten aufzuzeigen, wie man diese Erfahrungen entweder allein oder gemeinsam mit anderen verarbeiten kann.

⁰² Cope

Während sich Heal auf die emotionale Bewältigung konzentriert, liegt der Fokus bei Cope stärker auf präventiven Maßnahmen. Hier erhalten Nutzer:innen Unterstützung dabei, wie sie mit ähnlichen Situationen in Zukunft umgehen können und welche Beratungsstellen oder Anlaufstellen für ihre individuellen Bedürfnisse passend sein könnten.

⁰³ Fight back

Was kann man tun, um sich aktiv gegen rassistisch motiviertes Verhalten zu wehren? Dieser Bereich befasst sich mit rechtlichen und interpersonellen Strategien, um Betroffenen konkrete Möglichkeiten aufzuzeigen, wie sie sich juristisch oder gesellschaftlich gegen Diskriminierung zur Wehr setzen können. Dabei wird eine kompakte und übersichtliche Zusammenstellung relevanter Informationen bereitgestellt.

Mit der Zeit wurde deutlich, dass die anfänglich definierten Topics zwar gute Orientierungspunkte waren, aber nicht ausreichend waren, um die Vielfalt der individuellen Erfahrungen abzubilden. Jeder einzelne Fall bringt seine eigene Tiefe und Dringlichkeit mit sich – eine Herausforderung, die ein reines NLP-basiertes Modell nicht bewältigen konnte. Um dem gerecht zu werden, wurde ein vollständiger LLM-basierter Ansatz als der geeignetere Weg identifiziert.

→ Zentrale Erkenntnisse

- ✓ Nutzer:innen wollten **freier formulieren** und ihre Erfahrungen detaillierter teilen.
- ✓ Der regelbasierte Ansatz war **zu unflexibel**, da er nicht auf individuelle Bedürfnisse eingehen konnte.
- ✓ Ein **LLM-basierter Ansatz** könnte helfen, tiefere und natürlichere Gespräche zu ermöglichen.

HYBRID

Nach dem MVP wurde 2023, nach einer intensiven Re-Design-Phase eine neue

hybride Version gelauncht – eine Kombination aus NLU, Buttons und einer LLM-Komponente.

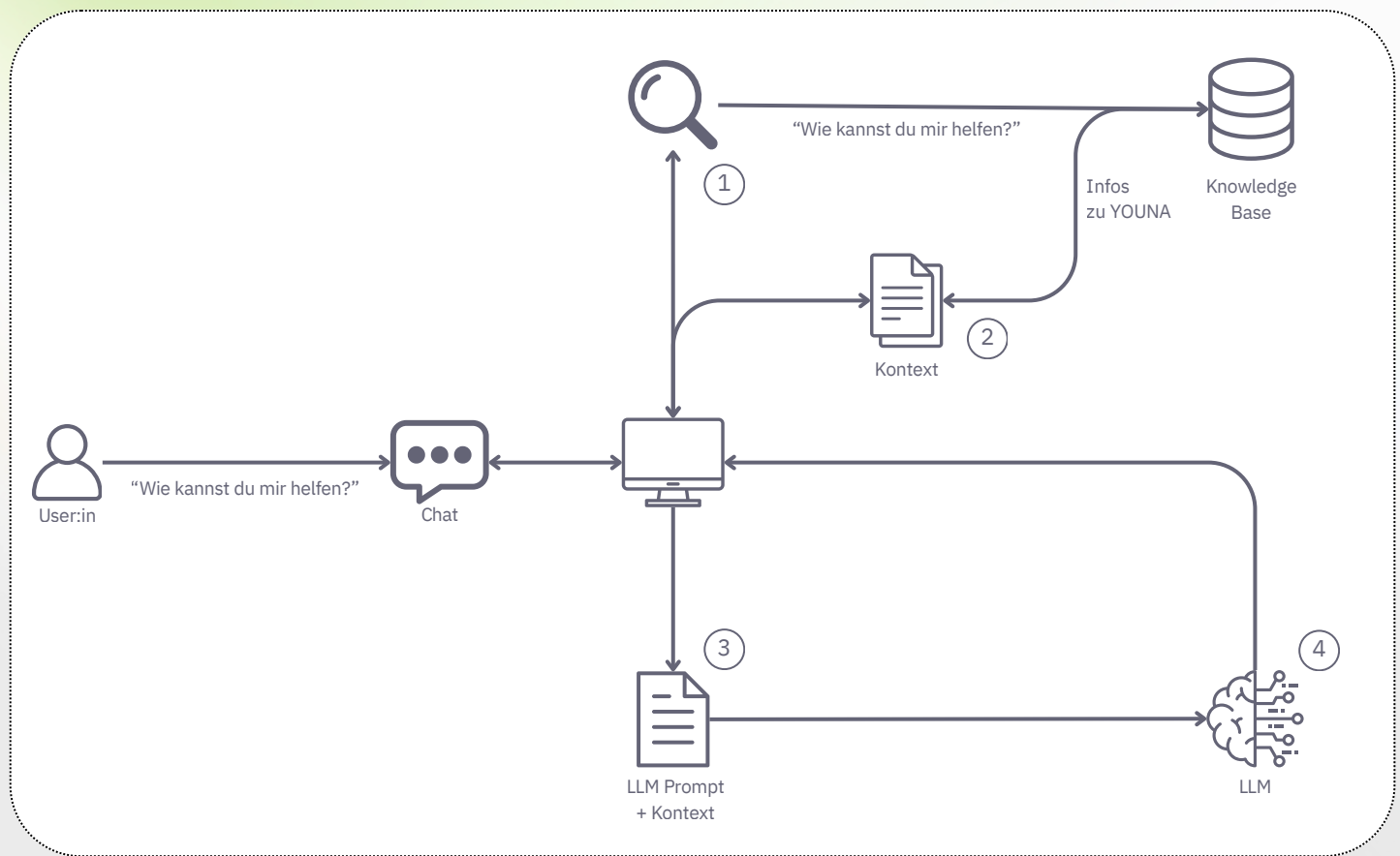
Wir implementierten einerseits einen Sub-Dialog, in welchem die User zum ersten mal selbst frei von ihren Erfahrungen berichten konnten. Allerdings blieb die Interaktion stark eingeschränkt durch vorgegebene Folgefragen des Bots. Der offene Dialogpfad war auch nach wenigen Interaktionen wieder vorbei. Der Fokus lag primär auf Zuhören, nicht auf echter inhaltlicher Tiefe. Ein nächster Schritt in Richtung dynamischen Modell war ein weiteres LLM, welches während des gesamten Gesprächsverlauf darauf achtete, dass die vorgegebenen Antworten auch zum Kontext passen.

Während der Hybridphase wurde klar, dass ein klassischer Chatbot oft nicht den Bedürfnissen der Nutzer:innen entspricht. Zu Beginn wussten wir nicht genau, welche Fragen gestellt werden würden oder welche Unterstützung wirklich benötigt wurde. Mit dem LLM ist es nun möglich, das gesamte gesammelte Wissen kontextsensitiv zu nutzen und den User:innen eine individuellere Unterstützung zu bieten - ein entscheidender Vorteil gegenüber dem regelbasierten Ansatz.

Schritt für Schritt wurden neue Ansätze getestet, um den Übergang von diesem geschlossenen Modell hin zu einem offenen Antwortformat zu ermöglichen. Die Entwicklung von YANA zu YOUNA war daher ein kontinuierlicher Prozess aus Development, Testing und Re-Design, um den Anforderungen der Nutzer:innen noch besser gerecht zu werden. Zwar kam die hybride Version unseren Vorstellungen bereits näher, doch um wirklich auf individuelle Bedürfnisse eingehen zu können, war ein weiterer Schritt nötig.

Die besondere Herausforderung unseres hybriden Ansatzes bestand darin, dass für die NLU nach wie vor eine große Menge an Trainingsdaten notwendig war, um einen kohärenten Gesprächsverlauf zu gewährleisten. Da diese Voraussetzung für uns nicht realisierbar war, musste eine alternative Lösung gefunden werden.

2024 fiel schließlich die Entscheidung, den hybriden Ansatz hinter uns zu lassen und vollständig auf ein LLM zu setzen, welches wir anhand einer dynamisch erweiterbaren, selbst kuratierten Knowledgebase tunen – ein Wechsel, der nicht nur das Dialogsystem grundlegend verbesserte, sondern auch unser eigenes Aufgabenfeld deutlich erweiterte. YOUNA was born.



YOUNA Um den komplexen Charakter von YANA beizubehalten und gleichzeitig die Möglichkeiten eines Large Language Models (LLM) zu nutzen, wurde ein neuer Ansatz gewählt. Dabei wird ein vorhandenes LLM – in diesem Fall GPT-4o – gezielt darauf optimiert, seine Antworten nicht nur aus den ursprünglichen Trainingsdaten zu generieren, sondern auf eine erweiterbare externe Wissensquelle zurückzugreifen.

Für die Umsetzung wurde eine Knowledge Base von YANA erstellt – das Äquivalent zu einem menschlichen Wissensgedächtnis. Diese umfasst sowohl die ursprünglichen Inhalte, auf denen YANAs Antworten basierten, als auch zusätzliche, tiefergehende Informationen zu komplexen Themen.

Die Flexibilität der Knowledge Base-Integration zeigt sich besonders in der Anpassungsfähigkeit des Systems. Durch eine kontinuierliche Analyse der Gesprächsverläufe konnten wir feststellen, dass viele Nutzer:innen gezielt nach spezifischen Beratungszentren fragten.

Ursprünglich war dieses Thema nur oberflächlich in der Datenquelle vorhanden – durch die neue Struktur ließ es sich jedoch mit minimalem Aufwand erweitern, um tiefere Informationen bereitzustellen. Zusätzlich wurde das Empathiepotenzial des LLMs gezielt durch Prompt Engineering verstärkt.

Vorteile des neuen Ansatz:

- ✓ Längere und differenziertere Dialoge sind möglich
- ✓ Besseres Verständnis für Nutzer:innen-Bedürfnisse durch die Analyse häufiger Themenanfragen
- ✓ **Einfache Erweiterung und Spezialisierung** der Knowledge Base je nach Bedarf

Der Fokus lag darauf, die Sensibilität im Umgang mit traumatischen Erfahrungen weiterzuentwickeln und sicherzustellen, dass der Bot nicht nur sachlich antwortet, sondern auch auf emotionaler Ebene eingeht. Dieser Bereich wird kontinuierlich optimiert, um die Interaktionsqualität zu verbessern.

No-Go

No-Go's jeder Art. Das können etwa Antworten sein, welche als beleidigend, rassistisch oder diskriminierend empfunden werden. Jedoch fallen in diese Kategorie auch jegliche unverständliche oder zu kompliziert formulierte LLM-Ausgaben. Auch, wenn es an Empathie fehlt oder man sich beim Lesen vor den Kopf gestoßen fühlt.

Fact Check

Hierbei betrifft es vor allem Antworten, die sich mit Beratungsstellen, Organisationen oder generellen Ressourcen befassen. Es soll geprüft werden, ob diese faktisch korrekt sind, existieren und im gegebenen Kontext auch empfehlenswert sind.

Hilfe zur Selbsthilfe

Es geht es darum, inwiefern der Bot Unterstützung bietet, ohne belehrend oder aufdringlich zu wirken. Es sollten dabei vorschnelle oder nicht hilfreiche Ratschläge vermieden werden, vor allem, wenn der/die Nutzer:in eher an emotionaler Unterstützung interessiert ist als eine Lösung zu suchen. Auch inwieweit vorgeschlagene Ressourcen zum Thema passen, ob diese auf den individuellen Kontext der Person abgestimmt sind und diese auch explizit gewünscht werden oder die Situation danach verlangt.

EVALUATION

Um dieses neue LLM-System kontinuierlich zu bewerten und zu verbessern, haben wir auf Langfuse zurückgegriffen - eine Open Source Engineering Plattform für LLM Analysen. Hierbei

lag im Vergleich zu einem Customer Service Bot die Challenge daran, dass es sich um ein subjektives Thema handelt und das Ziel, dem/der User:in vollständig zu helfen, nicht erfüllbar ist. Als Lösung haben wir die folgenden Custom Metrics eingeführt, welche kontinuierlich für die Gesprächsverläufe evaluiert werden.

Empathe

Hier wird danach bewertet, ob die Antwort Wörter und Formulierungen enthält, die Wärme, Respekt und Fürsorglichkeit vermitteln (z.B. „Ich bin für dich da“ statt „Alles wird gut“). Es sollen distanzierte oder sachlich-nüchterne Formulierungen bei emotional aufgeladenen Themen vermieden werden.

Relevanz / Nutzen

Der Chatbot bietet praktische Hilfestellungen oder stellt weiterführende Fragen, um mehr über das Erleben der benutzenden Person zu erfahren, falls es angemessen ist. Die Antwort ist an den Schweregrad der Emotion angepasst (z.B. vermeidet leichte oder humorvolle Antworten bei ernsten Themen).

Erkennen & Validieren von Emotionen

Die Antwort enthält Hinweise, die die Emotionen des Benutzers oder der Benutzerin explizit ansprechen (z.B. „Es klingt, als wärst du frustriert“ oder „Ich verstehe, dass das gerade schwierig für dich ist“) und/oder Formulierungen, die Verständnis ausdrücken (z.B. „Es ist völlig nachvollziehbar, dass du dich so fühlst“). Es werden Verharmlosungen oder unangebrachte Optimismusbekundungen vermieden, die die Emotionen der Person abwerten könnten. Auch wenn der Bot Hypothesen formuliert über die möglichen Ursachen / das zugrundeliegende universelle menschliche Bedürfnis hinter der Emotion des Nutzers oder der Nutzerin, wird das in dieser Kategorie analysiert.

RESULTS

Bei den Gesprächen, die YOUNA seit ihrer Einführung im Dezember geführt hat, konnte ein Anstieg der durchschnittlichen Gesprächslänge um mehr als 90% verzeichnet werden.

Im Durchschnitt konnten wir unsere Turns, also wie oft der Gesprächspartner und der Bot interagieren, um fast das 3 Fache steigern.

Insbesondere im Bereich Fact-Checking zeigt YOUNA bereits hervorragende Leistungen. Ein bekanntes Problem vieler LLMs ist die sogenannte Halluzination, die auch hier eine Herausforderung darstellt. Durch gezieltes Prompting und die kontinuierliche Erweiterung der Knowledgebase konnten wir diesem Problem jedoch effektiv entgegenwirken.

Hilfe zur Selbsthilfe ist eine weitere Stärke des Bots. Hier konnte beobachtet werden, wie Personen bei konkreten Problemen unterstützt wurden aus einer stets unterstützenden Position.

+90 %

Anstieg der
durchschnittlichen
Gesprächslänge

3x

so viele **Turns**



**Fact Checking,
Hilfe zur
Selbsthilfe &
Relevanz**

2x

so viele
Besucher:innen
senden eine
Nachricht

Eine ebenfalls erfolgreiche Steigerung konnten wir bei den Website Besucher:innen, die mindestens eine Nachricht senden, beobachten. So starten nun fast doppelt so viele Personen, welche die Website aufrufen, auch Konversation mit dem Chatbot.

Auch bei Nutzer:innen Befragungen konnten wir nach dem Umstieg auf einen vollständigen LLM Ansatz ein durchwegs positives Ergebnis erzielen. Dies spiegelt sich auch bei der Auswertung unserer Custom Metrics wieder.

Zudem haben unsere Custom Metrics gezeigt, dass ein Großteil der Antworten von YOUNA bereits eine hohe Relevanz aufweisen. Der Bot ist dabei in der Lage, auf verschiedene Sprachen einzugehen und auch anspruchsvolle Themen kritisch zu behandeln.

Fazit

Mit vergleichsweise geringem Aufwand im Gegensatz zu einem vollständig Voiceflow basiertem LLM konnten wir in kurzer Zeit mit unserem Ansatz einen leistungsfähigen, komplexen und vor allem empathischen Chatbot erstellen.

Learnings

Gerade für Anwendungsfälle, in denen nicht genau klar ist, was die User:innen fragen wollen und allgemein wenig Datenlage vorhanden ist, sind LLMs ideal. Im Gegensatz dazu sind regelbasierte Chatbots gut für vordefinierte Abläufe, sind aber nicht zu flexibel für offenere Gespräche.

Next Steps

- ✓ Weitere Optimierung der Knowledgebase, um auf neue Bedürfnisse einzugehen. Hier wollen wir uns vor allem auf den Ausbau von Supportgruppen & Netzwerken fokussieren, als auch - wie schon erwähnt - tiefer auf die Beratungsstellen eingehen. Ein weiteres Thema, auf welches wir tiefer eingehen wollen ist mental health im Bezug auf Rassismuserfahrungen.
- ✓ Durch gezielte Messung, welche wir anhand unserer Custom Metrics schon begonnen haben, wollen wir noch tieferes Verständnis für das LLM bekommen und wie wir es gezielter auf die Needs unserer User:innen anpassen können.
- ✓ Langfristige Weiterentwicklung mit Fokus auf Empathie, Zugänglichkeit und Effizienz. Dies wird vor allem durch unser Prompting verwirklicht und anhand von laufenden Befragungen und Testungen stetig verbessert.

Contact

Das beste zwischen Mensch und
Machine.

📍 Schönhauser Allee 141 B

📞 +49 30 555741390

✉ hi@lautmaler.de

🖱 www.lautmaler.de



Lautmaler